International Journal of Computer Science Engineering
and Information Technology Research (IJCSEITR)
ISSN(P): 2249-6831; ISSN(E): 2249-7943
Vol. 4, Issue 1, Feb 2014, 151-160
© TJPRC Pvt. Ltd.

ITERATIVE ALGORITHM FOR EXTRACTION AND DATA VISUALIZATION OF HL7 DATA

MOHANJIT KAUR[1], HARDEEP SINGH KANG[2] & K. S MAAN[3]

[1]Research Scholar, GNDEC Ludhiana, Punjab, India

[2]Assistant Professor, Department of Computer Science and Engineering, RIMT, Mandi Gobindgarh, Punjab, India

[3]Associate Professor, Department of Information and Technology, GNDEC, Ludhiana, Punjab, India

ABSTRACT

Health care has become one of the most important services. Hospitals, physicians, insurers, and managed-care firms are networking, merging and forming integrated organizations to finance and deliver health care. Hospitals, doctors, and other healthcare centers around the world require the ability to send and receive healthcare data, including patient information and various lab reports means that vast amounts of healthcare information are exchanged on a daily basis. However medical data can be extremely complicated due to the abundance of clinical terminology, as well as the structural complexity in the formation of the presented information

The objective of the present study is to extract useful information from the medical images stored in HL7 messages. In order to achieve this objective we first extracted images from HL7 meta data and messages and its base by using JAVA followed by data clustering using Multiple clustering algorithm which includes voltage , weak component and new proposed clustering and finally visualization of the data by creating graph diagrams based on graph theory. The results shows that based on certain criteria the dense connections in graphs can be reduced without the loss of information and in fact increased the visibility leading to production usage of information without clutter and noise in the presentation.

KEYWORDS: Health Care, Messages, HL7, Clustering

INTRODUCTION

Introduction to HL7

Hospitals and other healthcare provider organizations typically have many different computer systems used for everything from billing records to patient tracking. All of these systems should communicate with each other (or "interface") when they receive new information but not all do so. Health Level Seven International (HL7) founded in 1987, is a not-for-profit, ANSI-accredited standards developing organization dedicated to providing a comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of health services.

HL7, which is an abbreviation of Health Level Seven, is a standard for exchanging information between medical applications.

Health Level 7 (HL7) specifies a number of flexible standards, guidelines, and methodologies by which various healthcare systems can communicate with each other. HL7 provides standards for interoperability that improve care delivery, optimize workflow, reduce ambiguity and enhance knowledge transfer among all the stakeholders, including healthcare providers, government agencies, the vendor community, fellow SDOs and patients. Theoretically, this ability to

exchange information should help to minimize the tendency for medical care to be geographically isolated and highly variable.

HL7 Develops

- HL7 RIM, conceptual standards
- HL7 CDA, document standards
- HL7 CCOW, application standards and
- HL7 v2.x and v3.0, messaging standards. Messaging standards are particularly important because they define how information is packaged and communicated from one party to another.

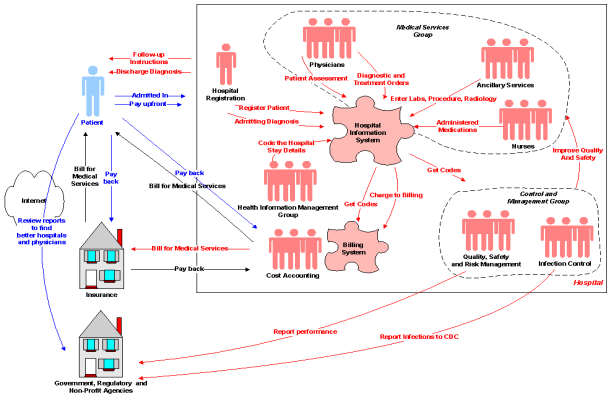

Figure 1: Working of HL7

In an HL7 message, the name of each segment in the message is specified by the first field of the segment, which is always three characters long. The following example message contains four segments: MSH, PID, NK1 and PV1. Different types of HL7 messages contain different segments. Here is an example of a typical HL7 message:

```
MSH|^~\&|EPIC|EPICADT|SMS|SMSADT|199912271408|CHARRIS|ADT^A04|1817457|D|2.5|
PID||0493575^^^2^ID 1|454721||DOE^JOHN^^^^|DOE^JOHN^^^^|19480203|M||B|254 MYSTREET
AVE^^MYTOWN^OH^44123^USA||(216)123-4567|||M|NON|400003403~1129086|
NK1||ROE^MARIE^^^^|SPO||(216)123-4567||EC|||||||||||||||
PV1||O|168 ~219~C~PMA^^^^^^^^^^|||277^ALLEN MYLASTNAME^BONNIE^^^^|||||||||
||2688684|||||||||||||||||||||199912271408||||0023 76853
```

Figure 2: HL-7 Message

The segments in this example contain the following information:

- The MSH (Message Header) segment contains information about the message itself. Every HL7 message specifies MSH as its first segment.
- The PID (Patient Information) segment contains demographic information about the patient, such as name, patient ID and address.
- The NK1 (Next of Kin) segment contains contact information for the patient's next of kin.
- The PV1 (Patient Visit) segment contains information about the patient's hospital stay, such as the assigned location and the referring doctor.

Over 120 different segments are available for use in HL7 messages.

The data contained in these messages can be used to extract meaningful information regarding types of patients being handled by a particular hospital or prevalence of a particular disease in a defined geographical area etc. The designed algorithm will be able to read and interpret multiple messages and present the information in a simplified format which can be readily understood.

Need of HL7

Clinical care facilities typically use a variety of complex software applications from different vendors. Because these applications are created by different software teams, these applications need to exchange data and typically do so via interfaces. An HL7 interface requires a sending and receiving module. These modules are created by the software vendor who programmed the application. In order to bridge the differences in HL7 format, modifications need to be made to the sending or receiving modules or an interface engine is used in the middle to translate the messages.

The presence of an HL7 interface engine in a healthcare environment gives more control to your organization and saves money and time by:

- Reducing the required number of export and import endpoints
- Allowing for reuse of data between applications
- Providing an easier method to interface a new or replaced application
- Providing the ability to monitor the entire system at one time
- Providing the ability to proactively notify interested persons using visual display and e-mail, when problems arise

Data Extraction

Data extraction is the act or process of retrieving data out of (usually unstructured or poorly structured) data sources for further data processing or data storage (data migration). The import into the intermediate extracting system is thus usually followed by data transformation and possibly the addition of metadata prior to export to another stage in the data work flow. Typical unstructured data sources include web pages, emails, documents, PDFs, scanned text, mainframe reports, spool files etc. Extracting data from these unstructured sources has grown into a considerable technical challenge whereas historically data extraction has had to deal with changes in physical hardware formats, the majority of current data extraction deals with extracting data from these unstructured data sources, and from different software formats.

The act of adding structure to unstructured data takes a number of forms

- Using text pattern matching such as regular expressions to identify small or large-scale structure e.g. records in a report and their associated data from headers and footers;
- Using a table-based approach to identify common sections within a limited domain e.g. in emailed resumes, identifying skills, previous work experience, qualifications etc. using a standard set of commonly used headings (these would differ from language to language), e.g. Education might be found under Education/Qualification/Courses;
- Using text analytics to attempt to understand the text and link it to other information

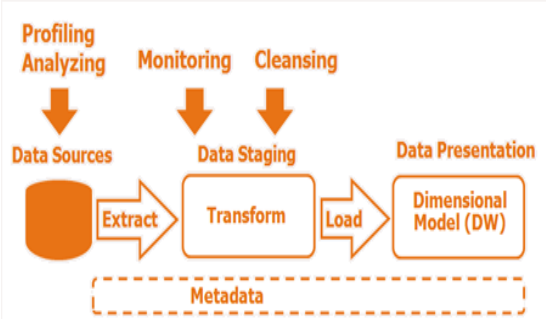

Figure 3: Data Extraction Process

PRESENT WORK

Problem Formulation

Any endeavor to develop an system for Data Visualization of HL7 data in medical area must follow International standards and integration possibilities. Since the treatment of any patient may be spread across various departments across many topologies and demographics. The information of a patient is the most critical resource for any medical fertinity to move ahead in curing and diagnosing a particular disease. For this HL7 standard was introduced for standardizing the sharing of critical patient's information and medical infrastructure. Therefore we purpose a new model which shall be developed on the system of patient information sharing based on Data Visualization of Hl 7 messages and data. Our attempt will be to help Medical fertinity to identify similar cases from shared resources of many hospitals.

Effective and meaningful communication between the information system from different vendors requires standards. The proper use of such standards protects investments, simplifies upgrade and replacement of equipments by avoiding vendor specific proprietary systems. The issue now is making sense of all those signals and finding stories in the stream . That's where visualizations come in. Whether you're dealing with a static graph or a real-time data wave, the act of seeing data unlocks much of its utility. In our research, we shall try to make visualization framework that will incorporate clustering based on features of medical images

Objectives

- Develop a representative repository of HL7 messages.
- Develop an iterative algorithm to extract exploratory and statistical information from HL7 repository.
- Develop graphical data visualization of the data extracted to gain insights into disease patterns and evaluate performance.

Present Method

This section explains the methodology of the study. The study was carried out in a systematic and sequential manner as depicted in the following steps:

Step 1: Extraction of data from HL7 messages data base through appropriate tagging and pattern matching.

Step 2: Developing a dataset of HL7 attributes related to the visits of the subjects to various departments for medical examination, post and pre check up.

Step 3: If data is cluttered then go to step 4. If data is not cluttered go to Step 5.

Step 4: Run clustering algorithm to remove noise and clutter from the data.

Step 5: Visualization of data to gain meaningful information

The schematic representation of the methodology is provided in the figure below:

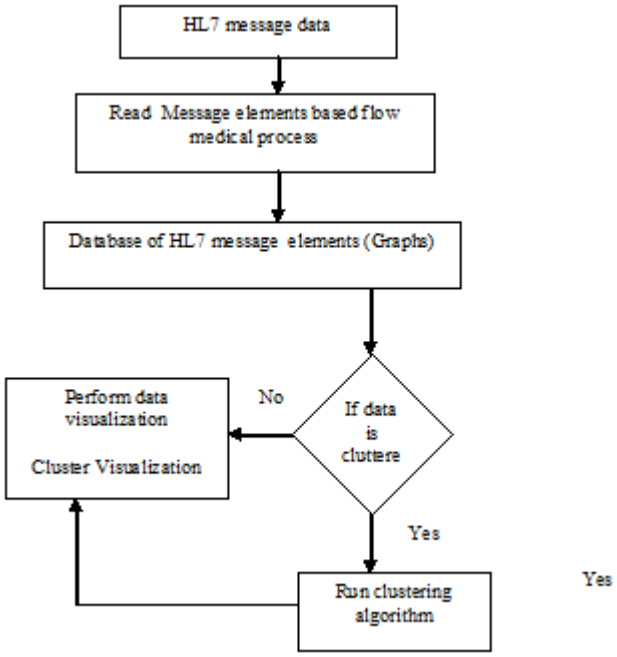

Figure 4: Schematic Representation of Methodology of the Study

Step 1: As a first in the study, a comprehensive dataset of HL7 messages elements to be extracted is identified. This dataset is to be used for further extraction of images to be studied

Step 2: The information from the above dataset are extracted by using element tags patterns. These tags help in identifying element values which are similar to each other in some aspect [weights /importance /precedence]. This will make the subsequent steps a more refined and results obtained will be more realistic and accurate.

```
MSH|^~\&|EPIC|EPICADT|SMS|SMSADT|          20091026120921          |CHARRIS|ADT^A04|1817457|D|2.5|
PID||0493575^^^2^ID          1|454721||DOE^JOHN^^^^|DOE^JOHN^^^^|19480203|M||B|254          MYSTREET
AVE^^MYTOWN^OH^44123^USA||(216)123-4567|||M|NON|400003403~1129086|
NK1||ROE^MARIE^^^^|SPO||(216)123-4567||EC|||||||||||||||||||
PV1||O|168          ~219~C~PMA^^^^^^^^^^|||277^ALLEN          MYLASTNAME^BONNIE^^^^|||||||||
||2688684|||||||||||||||||||||||199912271408||||||0023 76853
```

Figure 5: HL-7 Message Format

Read the above using regular expression and tagging we get the information about the customer. In an HL7 message, each segment of the message contains one specific category of information, such as patient information or patient visit data. For example, consider the HL7 message above, The segments in this example contain the following information:

- The MSH (Message Header) segment contains information about the message itself. This information includes the sender and receiver of the message, the type of message this is, and the date and time it was sent. Every HL7 message specifies MSH as its first segment.
- The PID (Patient Information) segment contains demographic information about the patient, such as name, patient ID and address.
- The NK1 (Next of Kin) segment contains contact information for the patient's next of kin.

- The PV1 (Patient Visit) segment contains information about the patient's hospital stay, such as the assigned location and the referring doctor

For Example

Table 1

Patient ID	Name of Patient	Date of Visit	Department	Location Distance
0493575	PID: DOE JOHN	20091026	ADT^A04 (Register a Patient)	PV134

Next, step is to build an adjacent matrix of the Patient as shown below:

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Figure 6: Adjacent Matrix

An adjacency matrix is a means of representing which vertices (or nodes) of a graph are adjacent to which other vertices. the adjacency matrix of a finite graph G on n vertices is the n × n matrix where the non-diagonal entry aij is the number of edges from vertex i to vertex j, and the diagonal entry aii, depending on the convention, is either once or twice the number of edges (loops) from vertex i to itself. Undirected graphs often use the latter convention of counting loops twice, whereas directed graphs typically use the former convention. There exists a unique adjacency matrix for each isomorphism class of graphs (up to permuting rows and columns), and it is not the adjacency matrix of any other isomorphism class of graphs. In the special case of a finite simple graph, the adjacency matrix is a (0,1)-matrix with zeros on its diagonal. If the graph is undirected, the adjacency matrix is symmetric.

Step 3) All the data collected in the step 2 is now stored as Graph Database , in fact , A graph database, also called a graph-oriented database, is a type of No SQL database that uses graph theory to store, map and query relationships.

Graph databases are well-suited for analyzing interconnections, which is why there has been a lot of interest in using graph databases to mine data from social media. Graph databases are also useful for working with data in business disciplines that involve complex relationships and dynamic schema, such as supply chain management, cargo transport and telecommunications .and Health network studies .The concept behind graphing a database is often credited to 18th century mathematician Leonhard Euler.

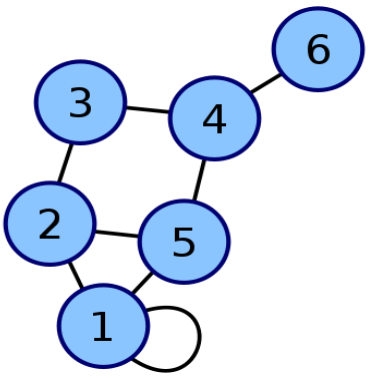

Figure 7

Where, these above is graph labels [1….6] representing the department visited by the patients.

Step 3: Once the HL7 elements have been initially extracted on the basis of medical process flow. If data so obtained is uncluttered the visualization is carried out to develop meaningful information. In case data so obtained is cluttered the same will be subjected to Bi-component, [Weak+Edge] and Proposed clustering (Step 4). The results obtained will be visualized using graph plots

Step 4: A comparison of the visualization graphs so obtained will be carried out before and after running the clustering algorithms.

CONCLUSIONS AND FUTURE SCOPE

Conclusions

Health care has become one of the most important services. Hospitals, physicians, insurers, and managed-care firms are networking, merging, and forming integrated organizations to finance and deliver health care. With the technological advancements, the application of computers has grown to a very large extent in almost every walk of life, especially the medical sciences. However, medical data can be extremely complicated due to the abundance of clinical terminology, as well as the structural complexity in the formation of the presented information. Thus, this information must be presented in a standardized format in order to ensure that the data is universally understood and organized.

The present study involves representation of the measures of the both the clustering algorithms (Bi component, Weak + Edge component clustering and Proposed algorithm), our intention was to identify particular process that can help to pack lots of numbers [of HL7 message artifacts data points] into a tiny space with very little distortion and its visualization maintains coherency, so that data can reveal more at micro and macro level to give a clear purpose to the cluttered dataset. Other than this our scheme of things was able to put the data in such a manner that it can be used for comparison between the different pieces in a easy manner with representative [un cluttered data] dataset‘s shape remaining close to the original dataset leading to high level of integration between the statistical and verbal descriptions of the HL 7 messages components.

Our process is able to get things done in minimal elapsed time between initial contact with the data and meaningful analysis. The time elapsed while using Volatile algorithm in case of 7 dimensional data is considerably less in case of graph visualization and first algorithm graph visualization plot, while it is slightly raised in case of Second [Edge +Week] cluster visualization. This effectively indicates that visualization using Proposed algorithm clustering is more enhanced in case of undirected of graph .as per of results that were obtained while running Proposed clustering.

This is incredibly important in current context of medical [HL7] environment because there are frequently demands for analysis within one day of the first contact with the data. A look at the visualization patterns will reveal that these make sense of the data with relatively little training, with meaningful patterns, trends, and easy to see and interpret HL7 message components trends. Moreover, it is pleasant to look at for long periods of time without undue visual fatigues.

Our visualization can answer real time Hl7 message components trends analysis medical questions. However, when it comes to interaction with the data (for example, to filter the data), it cannot do so in a manner that supports the flow of thoughts that might come about the data without interruption as it is based on simple direct input of the data to be visualized .

Future Scope

In the current research we have explored dataset of HL7 messages. Data represented in the form of graphs

network which have wide application by mapping the interaction of real life trends like the interaction of subjects /patients in seeking long term treatment from various departments of the hospitals. The collection of several data for doing time schedule analysis is worthwhile only if the data is represented and envisioned to give us proper insights /trends of the factors questioned however when this representation of graph becomes cluttered with overlapping data points then the purpose is defeated and therefore, we need frameworks of clustering as explored and built in current research which shows the visualization in terms of promising results, For future scope, we suggest that more algorithms of clustering may be explored with empirical applications in this problem area using these clustering and unsupervised algorithms which may include :

- Neural Network based Clustering
- Support Vector Based Clustering.

Other than the above directions for future, we would suggest that we must also work on other kinds of information that can be extracted from HL7 standard related protocol which are part of the overall hospital.

- Clinical Document Architecture
- Structured Product Labeling
- Electronic Health Record.

REFERENCES

1. Smith, B., & Ceusters, W. (2006). HL7 RIM: An Incoherent Standard. Studies in Health Technology and Informatics, 124, 133-138.
2. Custers, T., Arah, O.A., & Klazinga, N.S. (2007). Is there a business case for quality in The Netherlands?: A critical analysis of the recent reforms of the health care system. Health Policy.
3. Charles, N. (2006). Data Interchange Standards in Healthcare IT—Computable Semantic Interoperability: Now Possible but Still Difficult, Do We Really Need a Better Mousetrap?. Journal of healthcare information management: JHIM.
4. Dolin, R.H., Alschuler, L., Boyer, S., Beebe, C., Behlen, F.M., Biron, P.V., & Shabo, A. (2006). Hl7 clinical document architecture, release 2. Journal of the American Medical Informatics Association.
5. Shaver, D. (2002). Hl7 – What You Need to Know. Health Management Technology. Retrieved 14 September, 2007, from www.healthmgttech.com.
6. de Smet, A.G.M., & van de Klundert, J.L.M. (2005). Gegevensuitwisseling via het landelijk elektronisch medicatiedossier.
7. What is HL7 ?. (n.d.). Retrieved 1 November, 2007, from http://www.hl7.org/about/hl7about.htm#WhatisHL7
8. Vlug, A.E. (2005). Update HIS-tabel. Retrieved 25 Oktober, 2007, from http://www.raedelijn.nl/producten/innovatie/resolve Uid/713c59e5f301599bb3d6ab73272062bd
9. Klein, G.O. (2002). Standardization of Health Informatics – Results and Challenges. Methods of Information Medicine 41(4), 261-270.

10. NICTIZ. (2007). Beknopte versie Architectuurontwerp. Retrieved 17 Oktober, 2007, from http://www.aortarelease.nl
11. Smith, B., & Ceusters, W. (2006). HL7 RIM: An Incoherent Standard. Studies in Health Technology and Informatics, 124, 133-138.
12. Blobel, B.G.M.E., Engel, K., & Pharow, P. (2006). Semantic Interoperablility: Hl7 Version 3 Compared to Advanced Architecture Standards. Methods of Information in Medicine.
13. Branger, P. J., van't Hooft, A., van der Wouden, J. C., Moorman, P. W., & van Bemmel, J. H. (1999). Shared care for diabetes: supporting communication between primary and secondary care. International Journal of Medical Informatics.
14. Fernandez, E., & Sorgente, T. (2005). An analysis of modeling flaws in Hl7 and JAHIS. Proceedings of the ACM Symposium on Applied Computing 1, 216-223.
15. Klink, A. (2007). *Voortgang EPD*. Retrieved 20 Oktober, 2007, from

    http://www.minvws.nl/kamerstukken/meva/2007/vo ortgang-epd.asp